Received: 6 February 2023 | Accepted: 5 June 2023

DOI: 10.1111/2041-210X.14167

Methods in Ecology and Evolution

RESEARCH ARTICLE

A deep learning approach to photo–identification demonstrates high performance on two dozen cetacean species

Philip T. Patton[1,2] | Ted Cheeseman[3,4] | Kenshin Abe[5] | Taiki Yamaguchi[5] | Walter Reade[6] | Ken Southerland[4] | Addison Howard[6] | Erin M. Oleson[2] | Jason B. Allen[7] | Erin Ashe[8] | Aline Athayde[9] | Robin W. Baird[10] | Charla Basran[11] | Elsa Cabrera[12] | John Calambokidis[10] | Júlio Cardoso[9] | Emma L. Carroll[13] | Amina Cesario[14,15] | Barbara J. Cheney[16] | Enrico Corsi[10] | Jens Currie[1,17] | John W. Durban[18] | Erin A. Falcone[19] | Holly Fearnbach[18] | Kiirsten Flynn[10] | Trish Franklin[3,20] | Wally Franklin[3,20] | Bárbara Galletti Vernazzani[12,17] | Tilen Genov[21,22] | Marie Hill[2,23] | David R. Johnston[24] | Erin L. Keene[19] | Sabre D. Mahaffy[10] | Tamara L. McGuire[25] | Liah McPherson[1] | Catherine Meyer[26] | Robert Michaud[27] | Anastasia Miliou[28] | Dara N. Orbach[29] | Heidi C. Pearson[30] | Marianne H. Rasmussen[11] | William J. Rayment[31] | Caroline Rinaldi[32] | Renato Rinaldi[32] | Salvatore Siciliano[33] | Stephanie Stack[17,34] | Beatriz Tintore[28] | Leigh G. Torres[35] | Jared R. Towers[36] | Cameron Trotter[37] | Reny Tyson Moore[7] | Caroline R. Weir[38] | Rebecca Wellard[39,40] | Randall Wells[7] | Kymberly M. Yano[2,24] | Jochen R. Zaeschmar[41] | Lars Bejder[1,42]

Correspondence
Philip T. Patton
Email: pattonp@hawaii.edu

Funding information
National Oceanic and Atmospheric Administration; NOAA Fisheries QUEST Fellowship; University of Hawaii Information Technology Services; National Science Foundation, Grant/Award Number: 2232862 and 2201428

Handling Editor: Phil J Bouchet

Abstract

1. Researchers can investigate many aspects of animal ecology through noninvasive photo–identification. Photo–identification is becoming more efficient as matching individuals between photos is increasingly automated. However, the convolutional neural network models that have facilitated this change need many training images to generalize well. As a result, they have often been developed for individual species that meet this threshold. These single-species methods might underperform, as they ignore potential similarities in identifying characteristics and the photo–identification process among species.
2. In this paper, we introduce a multi-species photo–identification model based on a state-of-the-art method in human facial recognition, the ArcFace classification head. Our model uses two such heads to jointly classify species and identities, allowing species to share information and parameters within the network. As a demonstration, we trained this model with 50,796 images from 39 catalogues of 24 cetacean species, evaluating its predictive performance on 21,192 test images

from the same catalogues. We further evaluated its predictive performance with two external catalogues entirely composed of identities that the model did not see during training.

3. The model achieved a mean average precision (MAP) of 0.869 on the test set. Of these, 10 catalogues representing seven species achieved a MAP score over 0.95. For some species, there was notable variation in performance among catalogues, largely explained by variation in photo quality. Finally, the model appeared to generalize well, with the two external catalogues scoring similarly to their species' counterparts in the larger test set.
4. From our cetacean application, we provide a list of recommendations for potential users of this model, focusing on those with cetacean photo–identification catalogues. For example, users with high quality images of animals identified by dorsal nicks and notches should expect near optimal performance. Users can expect decreasing performance for catalogues with higher proportions of indistinct individuals or poor quality photos. Finally, we note that this model is currently freely available as code in a GitHub repository and as a graphical user interface, with additional functionality for collaborative data management, via Happywhale.com.

KEYWORDS
artificial intelligence, cetacean, computer vision, convolutional neural network, deep learning, dolphin, dorsal, lateral, machine learning, multi–species, photo–identification, whale

1 | INTRODUCTION

Many critical aspects of animal ecology, including movement (Palencia et al., 2021), demography (Borchers et al., 2014) and social behaviour (Bejder et al., 1998), can be efficiently studied by way of noninvasive photo–identification (photo–id). The photo–id process, starting from taking photographs of animals in the field to knowing the time and location of individual sightings, involves many resource–intensive steps. The matching step, that is, identifying the same individual in separate images, requires expertise and takes exponentially more time as the number of individuals in a catalogue grows. For example, Tyne et al. (2014) estimated that image matching for their one–year capture recapture survey of spinner dolphins *Stenella longirostris* cost over 1100 h of labour, nearly a third of the total financial costs for the project.

To mitigate these costs, researchers have developed several tools for automated matching. These tools are increasingly effective as more and more of them leverage recent advancements in deep learning (Bogucki et al., 2019; Clapham et al., 2020; Körschens et al., 2018; Miele et al., 2021; see Borowiec et al., 2022 for a thorough introduction to deep learning and its ecological applications). In this context, deep learning refers to training *convolutional neural networks* with many layers (definitions of *italicized* words can be found in the *Glossary*). While effective, these models often have millions of parameters (Tan & Le, 2019), making them prone to *overfitting*, and they struggle to identify individuals in images outside of the training set unless they are trained with many diverse images (Borowiec et al., 2022).

As a result, researchers have often focused on developing automated systems for individual species with enough images to train a neural network (Cheeseman et al., 2021; Clapham et al., 2020; Körschens et al., 2018; Maglietta et al., 2023; Miele et al., 2021; Thompson et al., 2021; but see Weideman et al., 2017). This single-species approach, however, has drawbacks. Photo–id surveys, such as camera traps, will often photograph any encountered species, even if the survey has a focal species. Additionally, this strategy passes over species with smaller training datasets, which is unfortunate because these rare or rarely observed species can be of conservation concern. Finally, single–species models ignore potential similarities in identifying characteristics and the photo–identification process among species. For example, species may have similar identifying marks such that a multi–species model would be able to transfer learning from one species to another (see the first two rows in Figure 1 for an example using cetaceans). Thus, a multi–species model that allows species to share information might outperform a single species model, particularly for species with few training images; this is a standard result in *transfer learning* (Zhuang et al., 2021).

In this paper, we introduce a multi–species approach to automated photo–id, then apply it to a large dataset of cetaceans, specifically, 39 catalogues of 24 species (see Supporting Information S1

FIGURE 1 Sixteen photo–id images of cetaceans, each row contains four images of the same individual, showing the challenges and opportunities for a multi–species photo–id model. The model must learn to recognize individuals from a variety of angles, under different lighting conditions, and in different social situations. Additionally, it must transfer learning from one species, for example false killer whale *Pseudorca crassidens* (top row), to a similar species, for example killer whale *Orcinus orca* (second row), while distinguishing them from less similar species like humpback whale *Megaptera novaeangliae* (third row) and very different species, like southern right whale *Eubalaena australis* (bottom row).

and S2 for a details on each catalogue). Photo–id forms the basis of nearly all capture–recapture modelling for cetaceans, which, in addition to distance sampling, is one of the primary methods for modelling demography of this taxon (Hammond et al., 2021). In addition, photo–id sheds light on movement ecology (Baird et al., 2008) and social structure (Bejder et al., 1998), which are crucial components for defining stock boundaries. The high labor costs associated with traditional image matching can inhibit these analyses, thereby hindering cetacean conservation.

Our objectives for this paper are three–fold. The first is to introduce a multi–species photo–id model based on a state–of–the–art method in human facial recognition, the ArcFace *classification head* (Deng et al., 2019, 2020; Ha et al., 2020). Our model uses two such heads to jointly classify species and identities, allowing species to share information via shared *weights* within the network. Secondly, we demonstrate this approach on a large, multi–species dataset of cetaceans. The dataset was assembled for a data science competition that challenged teams to identify individual cetaceans from images of their dorsal/lateral side. The winning team developed the model presented in this paper. We train and evaluate this model using the competition data, comprising 50,796 training and 27,944 test images. As a further evaluation step, we test the model on two additional catalogues that were not included in the competition dataset. Our final objective is to explain differences in performance among species and catalogues and provide recommendations to future users of the algorithm, specifically, to those with cetacean catalogues.

2 | MATERIALS AND METHODS

The objective for the multi–species photo–id model is to correctly predict the true identity of an animal, y_i, in the ith image regardless of species. This includes identities already in the catalogue as well as new individuals, which is known as an open–set recognition problem (Deng et al., 2019). That is, the model must be able to classify individuals outside of the training set as a "new individual." (sensu Maglietta et al., 2020; Miele et al., 2021).

2.1 | Multi–species photo–identification model

To facilitate discussion, we split the photo–id model into three parts: the *backbone*, the *neck* and the *classification heads*, that is the bottom row, in orange, of Figure 2. The figure shows the training pipeline of the model, specifically, the cetacean application, although other applications of this framework would likely look similar.

2.1.1 | The backbone and neck

The first step of the image through the network is into the *backbone*. Backbones are *convolutional neural networks* designed to be widely applicable to many problems in image classification, a broad category that includes photo–id. A flurry of research over the past

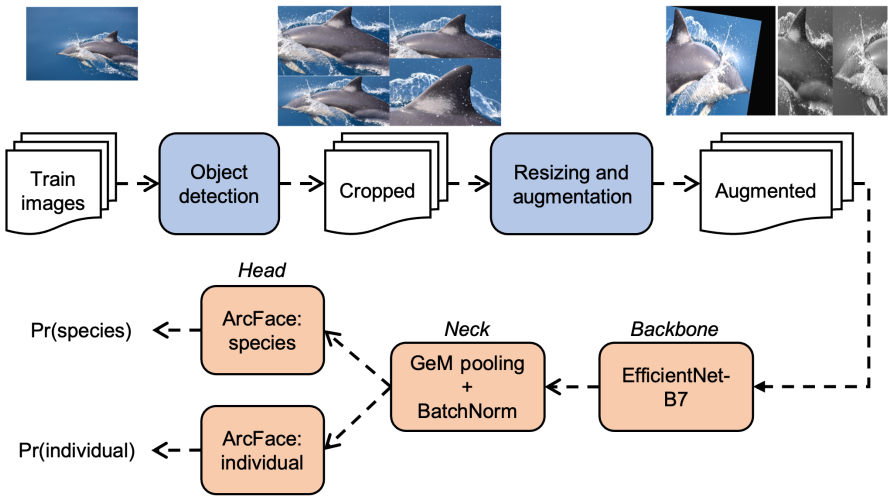

FIGURE 2 A model of the training pipeline for multi-species photo-id of cetaceans. The top row consists of preprocessing steps, shown in example by an image of a common dolphin *Delphinus delphis*. Crops generated by the four object detection models are shown, as well as two examples of images generated by the *data augmentation* step. The bottom row shows the training steps of the image classification network, from the backbone to the neck to the classification head. See Section 2 for a description of each component in the pipeline.

decade has produced dozens of popular backbones, including ResNet, DenseNet, Xception and MobileNet (Khan et al., 2020). *Validation* indicated that the EfficientNet-B7 backbone (Tan & Le, 2019) outperformed other similar backbones in the cetacean application.

The backbone takes an image and processes it through a series of *convolutional layers* and *pooling layers*, producing a reduced, three-dimensional representation of the image. The *neck* reduces this output to a one-dimensional vector, $\mathbf{x}_i$, where i is the image. This is known as the *feature vector* and, as we will see below, plays an important role in classification and prediction (Miele et al., 2021).

In the cetacean application, the neck builds the *feature vector* in two steps. First, it condenses the three-dimensional output of the backbone to a vector by way of generalized mean (GeM) pooling (Radenović et al., 2018), with $p = 3$, where p is the primary *hyperparameter* in GeM pooling. Next, it normalizes the vector with a batch normalization layer (Ioffe & Szegedy, 2015), which has been shown to improve the performance of the ArcFace classification heads (Deng et al., 2019, 2020; Ha et al., 2020).

2.1.2 | The classification heads

Our objective is to build a model that can classify individuals of multiple species. Further, the model should be able to *transfer learn* from one species to another with similar identifying characteristics. To that end, we use a model with two classification heads: one for species and one for individual (Figure 2). The input for both heads is $\mathbf{x}_i$, the output of the neck, meaning that both classification heads share *weights* in the backbone (Figure 2). These shared weights encourage the model to transfer learn between the two tasks—predicting species and identities—and, as it learns to predict species, it learns to better predict identities of those species. Thus, while we include every species in the same model, we do so in a way that acknowledges differences among species. If these differences are slight, then the model can readily transfer learning about one species to another, which is particularly helpful if one of these species has few training images and the other has many (Zhuang et al., 2021). Thus, this shared model, consisting of one backbone and two classification heads (Figure 2), accomplishes our goals of predicting identities regardless of species and encouraging transfer learning among species.

Both classification heads convert the feature vector, $\mathbf{x}_i$, to *class* probabilities, that is Pr(species) or Pr(individual) (Figure 2). Arguably, the most common method for computing these probabilities involves multiplying $\mathbf{x}_i$ by a matrix of weights, $\mathbf{W} \in \mathbb{R}^{D\times N}$, where N is the number of *classes* and D is the length of the feature vector (Deng et al., 2019). The resulting length-N vector of logits could be fed to the softmax function to get the N classification probabilities. This approach, coupled with the cross-entropy loss function, is sometimes referred to as softmax loss (Deng et al., 2019),

$$\ell_{\text{softmax}} = -\log \frac{e^{\mathbf{w}_{y_i}^T \mathbf{x}_i}}{\sum_{j=1}^{N} e^{\mathbf{w}_j^T \mathbf{x}_i}}, \quad (1)$$

where j indicates the class, y_i represents the identity in the ith image, and $\mathbf{w}_{y_i}$ are the weights associated with the y_ith identity.

Softmax loss tends to inadequately discriminate among classes in open-set photo-id, that is when the model must classify new identities in images after deployment (Deng et al., 2019). Other losses, such as triplet loss (Miele et al., 2021), are designed to correct this drawback. That said, triplet loss requires precise combinations of images in each training batch, significantly complicating and extending training when the dataset demonstrates extreme *class imbalance*, as is the case in many photo-id catalogues (Deng et al., 2019).

To rectify these issues, Deng et al. (2019) developed the Additive Angular Margin Loss function, otherwise known as ArcFace loss (Equation 2). ArcFace works by rewriting the logit in (Equation 1) as $\mathbf{w}_j^T \mathbf{x}_i = \| \mathbf{w}_j^T \| \, \| \mathbf{x}_i \| \cos\theta_j$, following the geometric definition of the dot product. Then, we fix the magnitude of the weight vector, $\| \mathbf{w}_j \|$, and the feature vector, $\| \mathbf{x}_i \|$, to 1 via l_2 normalization. Now, the loss incurred strictly depends on θ_j, the angle between the feature vector and the weights, which can be calculated as $\theta_j = \arccos\left(\mathbf{w}_j^T \mathbf{x}_i\right)$. Then, ArcFace adds a marginal penalty, m, to the angle θ_{y_i}. This is the angle between the feature vector $\mathbf{x}_i$

and the weights, $\mathbf{w}_{y_i}$, associated with the true class in the ith image. This penalty effectively pushes each feature vector closer to the centre associated with the true class, $\mathbf{w}_{y_i}$, and pulls it away from the centres of other classes, $\mathbf{w}_j$, (Figure S1; also see fig 3 in Deng et al., 2019). Then, we back-transform to logits by taking the cosine of the penalized angle, then re-scale the logits with a parameter s. Finally, the logits are fed into the softmax function to calculate the probabilities, which are evaluated with the target values using cross entropy loss. The final loss function can be written as

$$\ell_{\text{ArcFace}} = -\log\frac{e^{s\cdot\cos\left(\theta_{y_i}+m\right)}}{e^{s\cdot\cos\left(\theta_{y_i}+m\right)}+\sum_{j=1,j\neq y_i}^{N} e^{s\cdot\cos\theta_j}}. \quad (2)$$

For all its benefits, ArcFace loss will struggle when samples naturally lie far away from their class centres, $\mathbf{w}_{y_i}$ (Deng et al., 2020). This could happen if two images of the same individual show opposite flanks, the individual acquires new identifying features, or the existing marks are obscured. Under these circumstances, ArcFace loss may perform better if we allow each class to have multiple centres. This is known as Sub-center ArcFace (Deng et al., 2020) and works by adding an additional dimension to the weight matrix, $\mathbf{W}\in\mathbb{R}^{N\times K\times D}$, where K is the number of sub-centres per class. The loss function becomes

$$\ell_{\text{ArcFace}_{\text{subcenter}}} = -\log\frac{e^{s\cdot\cos\left(\theta_{i,y_i}+m\right)}}{e^{s\cdot\cos\left(\theta_{i,y_i}+m\right)}+\sum_{j=1,j\neq y_i}^{n} e^{s\cdot\cos\theta_{i,j}}}, \quad (3)$$

where $\theta_{i,j}$ is the angle between the feature vector and the sub-centre, which can be calculated as $\theta_{i,j}=\arccos\left(\max_k\left(\mathbf{W}_{j_k}^T\mathbf{x}_i\right)\right)$ for $k\in\{1,\ \ldots\ ,K\}$. The dot product, $\mathbf{W}_{j_k}^T\mathbf{x}_i$, produces a matrix, $\mathbf{S}\in\mathbb{R}^{N\times K}$, which can be viewed as cosine similarity scores between the feature vector and the sub-classes. The $\max_k\left(\mathbf{W}_{j_k}^T\mathbf{x}_i\right)$ effectively max pools the matrix $\mathbf{S}$ to produce a vector of N similarity scores between the feature and the class. This max pooling step, along with the additional centres for each class, helps alleviate the issue of mark changes or images from different angles. In the cetacean application, we set $K=2$ for both classification heads.

The final nuance of the classification heads involves the margin parameter, m, in (Equation 3). Learning the optimal value for the m can be challenging with heavily *imbalanced classes* (Ha et al., 2020). To address this issue, we let the value of the m vary by the number of images per class, $m_c=az_c^{-\lambda}+b$, where m_c is the margin value for class c, z_c is the number of images of the class, a is a coefficient, b is an intercept, and λ is the rate of decay to b as z grows. Thus, classes with fewer images, which tend to be harder to learn, have higher margins (Ha et al., 2020). In the cetacean application, we use this technique for both classification heads.

Altogether, these types of classification heads are known as Sub-center ArcFace with Dynamic Margins and have become popular in image search problems (Ha et al., 2020). As such, they should be generally applicable to many problems in multi-species photo–id, which are analogous to image search.

2.2 | Cetacean application

This approach was originally developed by the winning team, "Preferred Dolphin" of a Kaggle competition that lasted from February 1 to April 18, 2022. Kaggle is a platform that allows organizations to solve machine learning and data science problems via open–source competition. Its competitions have previously produced effective photo–id models (Bogucki et al., 2019; Cheeseman et al., 2021). In this case, Happywhale, a platform for recognizing individual humpback whales in photographs and sharing data among researchers and the public, challenged competitors to develop a model for recognizing any individual cetacean from a photograph of its dorsal side. 1588 teams competed for $25,000, submitting 39,284 predictions of individuals in the test set. Full competition details, including data, discussions, notebooks of code and the leaderboard, can be found at https://www.kaggle.com/competitions/happy-whale-and-dolphin.

The objective for the competition was to maximize the mean average precision (MAP) of the test set, specifically, the true identity of the individual in each test image. The competitors were to submit five predictions per image that were scored using precision. For a set of five ordered predictions, the precision score will be 1/1 = 1 if the first prediction is correct, 1/2 if the second is correct, and so on until 1/5 if the fifth prediction is correct, or 0 if none of the five predictions are correct. MAP is the mean precision score for a set. Teams submitted their predictions to Kaggle, which then reported the MAP score. This score was calculated using classified subset of the test set, that is the "public" set (6752 images). At the end of the competition, teams were ranked by their score on the remaining test images, that is the "private" set (21,192 images). This splitting makes for better model evaluation because it reduces the chance that teams will overfit to the test set. The competition's winning team, *Preferred Dolphin*, developed the model discussed in this paper.

2.2.1 | Training data

Happywhale and Kaggle coordinated with researchers across the globe to assemble a large, multi–species dataset of individual cetaceans, consisting of 41 catalogues of 25 different species. Two of the competition catalogues were omitted from this analysis because one consisted of 26 low-quality images for both training and test, while the other lacked a test set (Supporting Information S1 and S2). Each catalogue contained a single species. Ten species were represented in more than one catalogue, and four species were represented in more than two. The dataset demonstrated extreme class imbalance for both species (Figure 6, top panel) and individual identities. The 50,796 training images contained 15,546 identities. Of these identities, 9240 (59%) had only one training image, while 14,210 (91%) had five or fewer. Each image was assigned one label, that is, there was one "true" individual per image. For images with multiple individuals, the label corresponded to the most discernible individual in the photograph. The test/train split was determined using catalogue characteristics and thus varied by catalogue (top panel, Figure 6).

Many of the competition images had considerable background noise that is irrelevant to matching (Figure 3). To address this, several competitors trained models that automatically detected a cetacean in an image and drew a bounding box around it, permitting automated image cropping following the approach of Bogucki et al. (2019). The training pipeline (Figure 2) incorporated four of these detectors (Figure 3). The diversity of detectors added some robustness to cropping errors (see the grey whale panel in Figure 3), and can be viewed as a form of *data augmentation*. To incorporate these four detectors, we randomly choose one to crop each image during each training *epoch* (Figures 2 and 3). A dorsal fin only detector was selected with probability 0.15. One of three full body detectors was selected with probability 0.6, 0.15 and 0.05, respectively. These three detectors produced slightly different crops (Figure 3), and differed in architecture, with two using YOLOv5 (Redmon et al., 2015) and one using Detic (Zhou et al., 2022). Finally, the original, uncropped image was selected with probability 0.05. Links to descriptions of each detector, including code, can be found at input/README.md in the paper's GitHub repository, as well as in the discussion forum for the competition (see the link in *Cetacean application*).

After cropping, we resized each image to 1024 pixels by 1024 pixels to be compatible with the EfficientNet-B7 backbone. After resizing, we applied several *data augmentations*, including affine transformation, resize and crop, grayscale, gaussian blur, gaussian noise, grid shuffle, posterize, brightness and contrast, cutout, snow, rain and horizontal flip, all from the Albumentations library (Table S1). Data augmentations are crucial for this architecture, which achieves 100% accuracy for identity in training yet poor performance in *validation* without them, suggesting severe overfitting.

2.3 | Hyperparameter optimization and training

The classification heads in Equation 3 are highly sensitive to the *hyperparameters*, a, b, λ and s. To optimize these values for both heads, we used Optuna (Akiba et al., 2019). We did so with a smaller backbone, EfficientNet B0 and smaller images, (256 by 256 pixels). As with all the hyperparameter optimization here, we validated changes in the hyperparameters on a hold-out set representing one fifth of the training images. The optimal values were $s = 20.9588$ for the classification head for the individual identity and $s = 33.1383$ for the classification head for species. See Figure 4 for the optimal margin parameter values.

After setting values for the hyperparameters, we trained the model using larger images (1024 by 1024), an EfficientNet-B7 backbone, and an AdamW optimizer (Loshchilov & Hutter, 2017). We set the maximum *learning rate* for the backbone to $1.6e-3$, 10 times lower than the learning rate for the head $1.6e-2$ (Figure 5). Further, we used a learning rate scheduler with a linear warmup for six *epochs*, then cosine annealing with a $1.0e-2$ decay rate (sensu Loshchilov & Hutter, 2016, without restarts). We trained the model for 30 total epochs using every training image (Figure 5). The total training time took about 64 h on a Nvidia Tesla v100 GPU.

We also conducted two rounds of pseudo-labelling. Pseudo-labelling is a semi-supervised learning technique for adding images to the training set. To do so, we trained the model then predicted individual identities for the 27,944 images in the test dataset. Images whose predicted class probability was over 0.6 were added to the training set. Then, we retrained the model, repeating the process once more. Essentially, this process accepts the model's most confident predictions as true labels, which can be beneficial with heavily imbalanced data. In fact, these two rounds of pseudo-labelling increased the private test MAP by 0.016. On the other hand, pseudo-labelling risks overfitting to the images in the test and train sets, making it less likely that the model will perform well in production. As a result, it was essential to evaluate the model with not only the competition's test set but also on catalogues that were not used in training the model.

2.4 | Prediction and evaluation

To generate predictions for pseudo–labelling and the final evaluation, we used the metric learning approach (Miele et al., 2021) of estimating similarity scores between feature vectors of the training set to that of the predicted image. Specifically, we found the 500 nearest neighbours in the training set to the prediction image using cosine similarity as the distance metric. If the neighbours included multiple training images of the same individual, we took the maximum value of the similarity score of that individual. Finally, we added a dummy class for new individual, giving it a similarity score of 0.465, then sorted the similarity values to get the final predicted classes. Thus, if the six nearest images in the training set to the prediction image had {score:class} {0.9:A, 0.8:A, 0.5:B, 0.3C, 0.2:D, 0.1:E}, the predictions would be {0.9:A, 0.5:B, 0.465:new, 0.3:C, 0.2:D}. We evaluated these predictions on the 21,192 images in the private test set, excluding the public test images for better model evaluation. Note that we excluded 52 images of grey whale flukes from the evaluation since they are too unlike any other image in training and test sets; they do not show the dorsal/lateral view of the animal.

We also evaluated the model's predictive performance on two catalogues that were not included in the competition and entirely comprised of identities that the model had not seen during training. One is a catalogue of 4670 rough-toothed dolphin *Steno bredanensis* images from the main Hawaiian Islands, produced by Cascadia Research Collective, spanning 20 years of data collection (Baird et al., 2008). The other is a catalogue of 754 spinner dolphin images from Oʻahu, Hawaiʻi that was collected from 2020 to 2022 by the Marine Mammal Research Program at the Hawaiʻi Institute of Marine Biology. This catalogue only included highly and moderately distinct individuals, as well as excellent and good photos (sensu Rosel et al., 2011).

3 | RESULTS

The model's predictions for the competition test set of 21,192 images from 39 catalogues of 24 species attained a MAP score of 0.869. The precision varied among species (Figure 6, Table S2) and did not correlate with the number of training images or test images. The model was generally better at recognizing toothed whales than

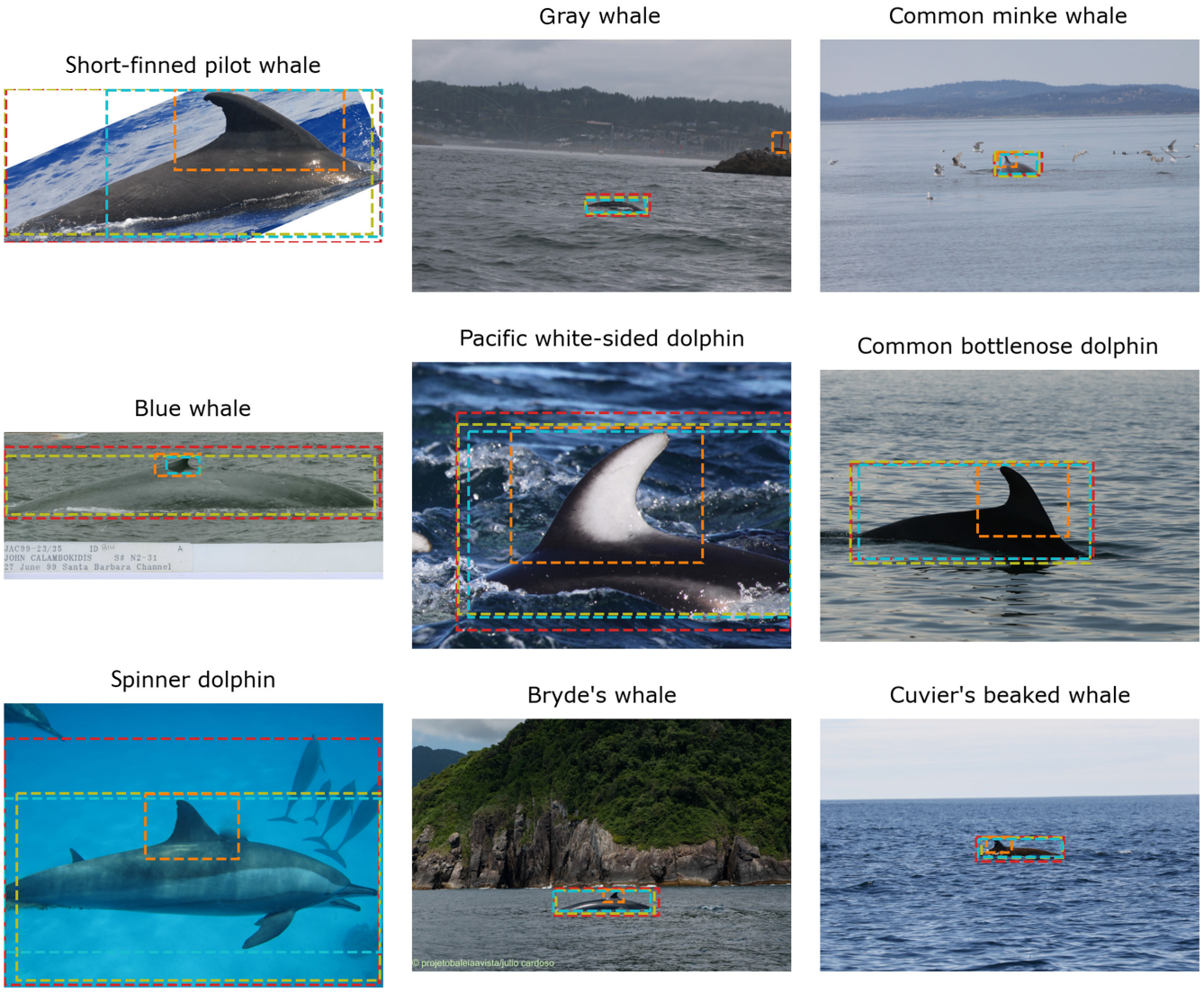

FIGURE 3 An image from nine catalogues in the competition set, and bounding boxes generated by the four cetacean detectors used by the model. The probability of seeing the crop generated by each bounding box was 0.60 for red, 0.15 for olive green, 0.15 for orange, and 0.05 for blue.

baleen whales; only two of the eight baleen whale species scored above average. These two species, sei whale *Balaenoptera borealis* and Bryde's whale *Balaenoptera brydei*, are primarily identified by their dorsal fins, as are most toothed whale species. The best performing species included so-called blackfish—false killer whale, melon-headed whale *Peponocephala electra*, pygmy killer whale *Feresa attenuata*, long-finned pilot whale *Globicephala melas*, short-finned pilot whale *Globicephala macrorhynchus*, and killer whale *Orcinus orca*—as well as common bottlenose dolphin *Tursiops truncatus* and spinner dolphin. The model was least able to recognize southern right whale and Cuvier's beaked whale *Ziphius cavirostris*.

For species with multiple catalogues, some exhibited significant variation in performance. For example, one catalogue of common minke whale *Balaenoptera acutorostrata* achieved 0.79 MAP while the other achieved 0.60. Other species with large disparities in performance among catalogues included beluga whales *Delphinapterus leucas* and killer whales. Conversely, humpback whales, bottlenose dolphins and spinner dolphins demonstrated consistent performance across catalogues. We were unable to find a metric to explain variation among catalogues. There was no consistent relationship between the catalogue-level MAP and mean image width, mean bounding box width, the number of training images, the number of distinct individuals, or the number of training images per individual (Figures 6 and 7). On the other hand, we found several qualitative measures that did appear to affect an image's precision score, including blur, distinctiveness, mark obfuscation, distance, contrast and splashing/spray (Figure 8).

The model was able to recognize individuals in the two unseen catalogues—the spinner dolphin and rough-toothed dolphin catalogues—better than these species' catalogues in the competition set. For example, the three competition spinner dolphin catalogues scored 0.96, 0.95 and 0.94, while the unseen catalogue scored 0.99. Similarly, the rough-toothed dolphin catalogue in the competition scored 0.74, while the unseen catalogue scored 0.83.

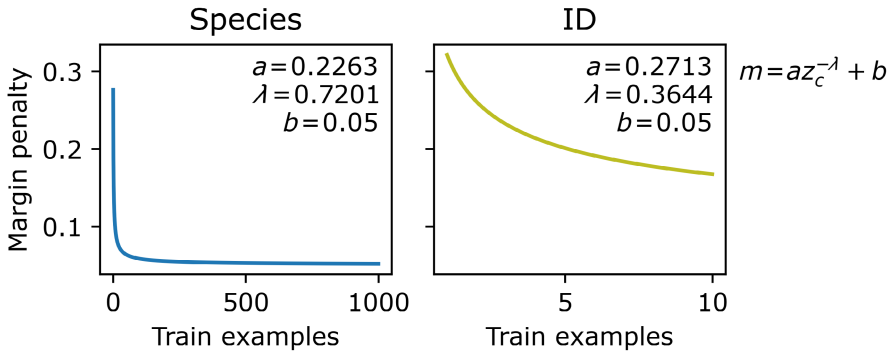

FIGURE 4 Dynamic margin curves for the classification heads for individual identity (ID) and species. The *y* axis is the margin penalty, that is, *m* in (Equation 3). The *x* axis represents the number training images for a given class. Note the different scales of the *x* axis. The median number of images by species was 866; for ID, it was 1, that is one training image per identity. Thus, the margin for a given identity was much higher than that of a given species as would be expected since individual identity is a more challenging recognition task than species identity.

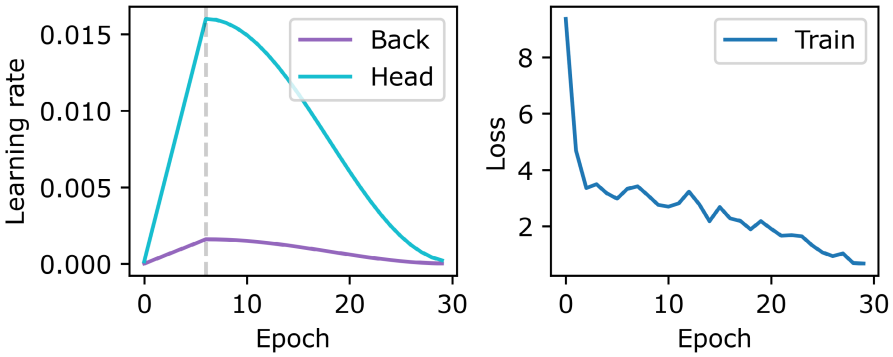

FIGURE 5 Learning rate scheduler and training curve. "Back," refers to the backbone and "Head," refers to the classification heads. The maximum learning rate for the backbone, indicated at the dashed line, was 10 times lower than that of the classification heads. The training curve shows the training loss over the epochs.

4 | DISCUSSION

Our multi–species approach to photo–id performed well in the cetacean application, scoring 0.869 MAP on the 21,192 test images. For reference, the humpback whale fluke model used by Happywhale.com achieved roughly 0.97 MAP in competition (Cheeseman et al., 2021) and, therefore, many humpback whale biologists trust the model to correctly identify the individual in nearly all circumstances. In our model, this level of precision (above 0.95) was attained by 10 catalogues representing seven species—false killer whale, killer whale, long–finned pilot whale, pygmy killer whale, common bottlenose dolphin, short–finned pilot whale and spinner dolphin (Figure 6, Table S2).

4.1 | Guidance for cetacean researchers

While identifying the causes of prediction errors in neural networks is challenging, we discuss possible explanations for the heterogeneity in performance among catalogues and species, and we make specific recommendations based on these findings. Further, for every example of a suspected cause of prediction failure in Figure 8, there are far more examples of the model accurately predicting an individual under the same conditions. That said, we have extensively studied the output, namely, the zero precision images, and have identified the following patterns.

4.1.1 | Species identified by their dorsal fin score best

All the highest performing catalogues represented species that are primarily identified by their dorsal fin. There are a few possible explanations for this pattern. Dorsal markings, for example, nicks and notches or distinctive curvature, are the primary identifying characteristic in 19 of the 24 species. Secondly, some of these species lack highly distinctive natural marks. As such, catalogues for these species, for example, the external spinner dolphin catalogue, often only contain individuals that acquire distinctive nicks and notches along the dorsal fin. Thus, there is a minimum level of distinctiveness for these catalogues, potentially making the model's job easier.

4.1.2 | Catalogues with fewer distinctive individuals will underperform

Catalogues without a distinctiveness threshold, for example the fin whale catalogue, may underperform since less distinctive animals are harder to identify in lower quality images (Rosel et al., 2011). Thus, the model's predictions for these individuals are more sensitive to nuisance characteristics in the image, likely lowering performance.

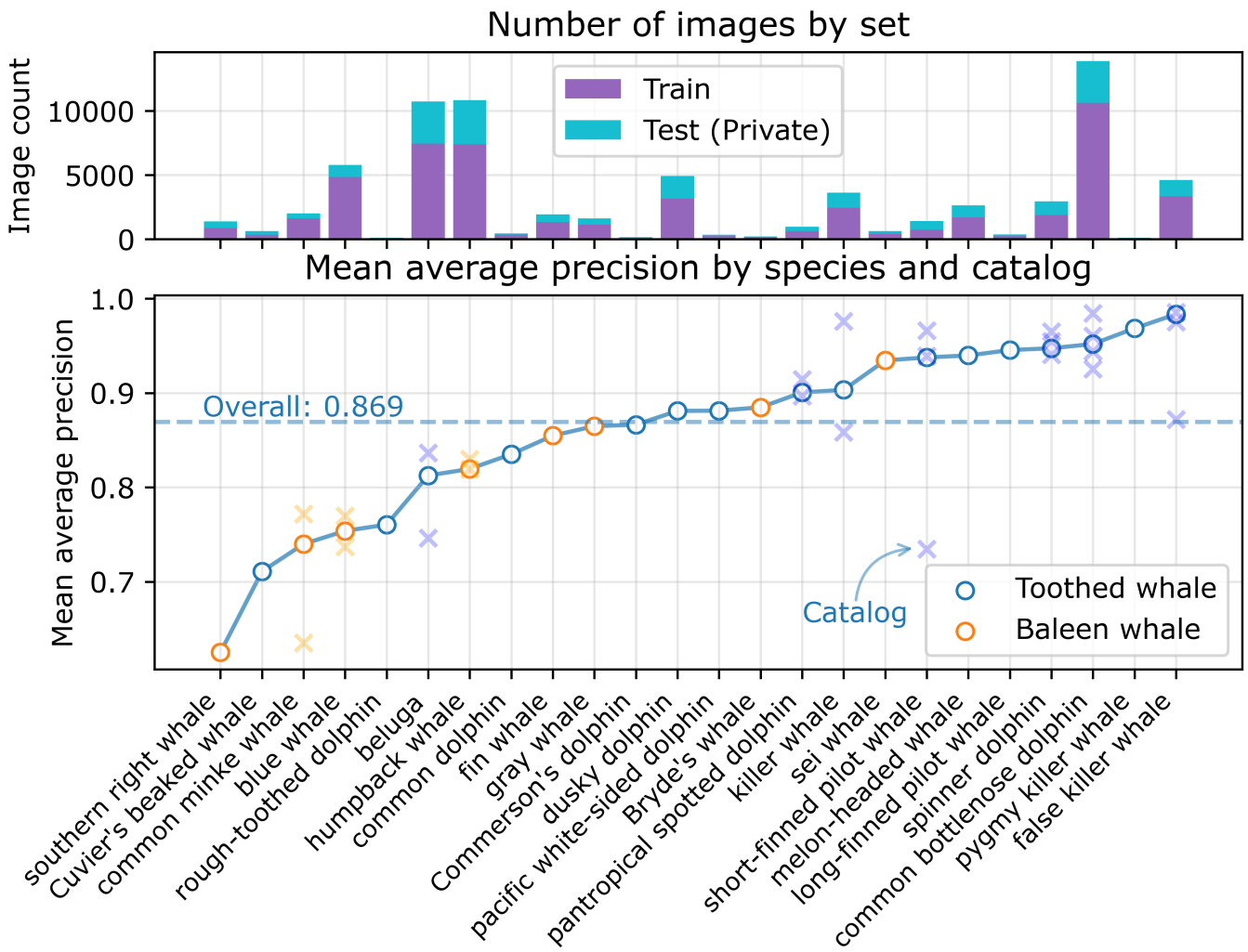

FIGURE 6 Mean average precision for the 24 species and 39 catalogues constituting the Kaggle test set. The top panel shows the number of images for each species by usage, that is, training or test. For species with multiple catalogues, the catalogue–level MAP is denoted by an *x*. The Fraser's dolphin catalogue, which scored 0.42 on 26 total images, is not pictured.

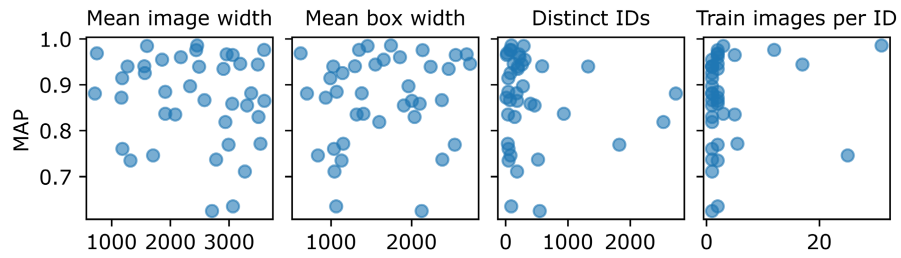

FIGURE 7 Possible explanatory variables for the catalogue level variation in performance. Each point is one of the catalogues in the competition dataset. Image and box width are represented in pixels. Box width is the weighted average of the bounding box, with weights corresponding to the probability of each box (Figure 3). Distinct IDs refers to the number of individuals in the training set. There was no consistent relationship between the catalogue–level MAP and mean image width, mean bounding box width, the number of training images, the number of distinct individuals or the number of training images per individual.

4.1.3 | Image quality is important

Most of the suspected causes of poor predictive performance relate to some measure of image quality. These include the angle of the subject to the camera, splash or spray covering the mark, blurred subject, the mark partially submerged in water, distance to the subject, and contrast of the mark against the background. Many of the differences in performance among catalogues for a given species could be attributed to average photograph quality. For example, the lower performing minke whale catalogue tended to have images taken from further away. Similarly, the lower performing beluga catalogue contained older, blurrier images than its higher performing counterpart.

FIGURE 8 Eight suspected causes of prediction errors, including six examples from the competition data. None of the predicted identities in each image were correct.

Some of these image quality measures relate to biology. For example, Cuvier's beaked whales are deep–divers that rarely spend time at the surface. As such, images of them are often taken from a distance. As another example, it is difficult to capture images of common dolphins without blur, breaches or splashes because they are fast and acrobatic swimmers.

4.1.4 | Pigmented animals might be harder to identify

Some species where pigmentation is used as a secondary identifying characteristic appeared to underperform. These included rough-toothed dolphin, pantropical spotted dolphin *Stenella attenuata*, Commerson's dolphin *Cephalorhynchus commersonii* and pacific white–sided dolphin *Lagenorhynchus obliquidens*. One possible explanation is that pigmentation for these species is not quite as distinctive as a mark such as, say, a large notch on the dorsal fin. Further, it may be difficult to capture a quality image of this pigmentation because of glare, contrast, splash/spray or other nuisance characteristics. Additionally, identifying pigmentation will vary from the left and right sides of the animals.

4.1.5 | Species with unusual marks relative to the training set will score poorly

This category primarily relates to the southern right whale, whose characteristic, callosities, may be too different from the other species in the

training data. While other high-performing catalogues had fewer images, for example pygmy killer whale, these catalogues often contained species with similar marking characteristics to other species with larger training sets, for example false killer whale. Conversely, there is no larger catalogue of a species with similar marks that can effectively transfer learn to the southern right whale catalogue.

4.1.6 | Preprocessing remains a hurdle

The model in Figure 2 incorporates a pre-processing step to automatically detect an animal and crop the image before feeding it to the photo-id model. Both this preprocessing model and the photo-id model were trained with the competition images, which all contained animals that had been identified to individual. Conversely, many raw images from the field will not contain animals that are identifiable because the images are poor quality, or the animal is indistinct. Therefore, our current preprocessing model likely would not work well for all raw field images. While the cetacean detection models used in the training pipeline (Figures 2 and 3) could be used to separate images with and without cetaceans, this model is not capable of removing images with cetaceans that are too blurry or indistinct for matching. Future research should focus on developing models for automating this grading step, which is essential to mark–recapture (Urian et al., 2015) and can be gruelling, such that raw field images can be processed before feeding this model.

4.1.7 | The model may struggle with mark changes

Mark changes may trick the model into classifying changed individuals as new, which could explain why the Antarctic killer whale catalogue performed worse than the Australian killer whale catalogue. Unlike some of their Australian counterparts, Antarctic killer whales tend to accumulate diatoms, which can completely obscure marks and may have impacted model performance. On the other hand, the structure of the model, that is the sub-centres in the classification head (Equation 3), may have made the model somewhat resilient to minor mark changes. Figure 9 shows one such example involving an individual bottlenose dolphin. In this example, six of seven test images show a new nick towards the top of the dorsal fin that was not present in the training images. Two of these were correctly classified. The other four were incorrectly classified as new individuals. For each of these images, however, the correct identity was included in the top five predictions. Thus, the model appears to show some resilience to a minor mark change. Nevertheless, future research should explore how to improve this resilience, perhaps by incorporating additional information, for example, social, spatial or other encounter-level data.

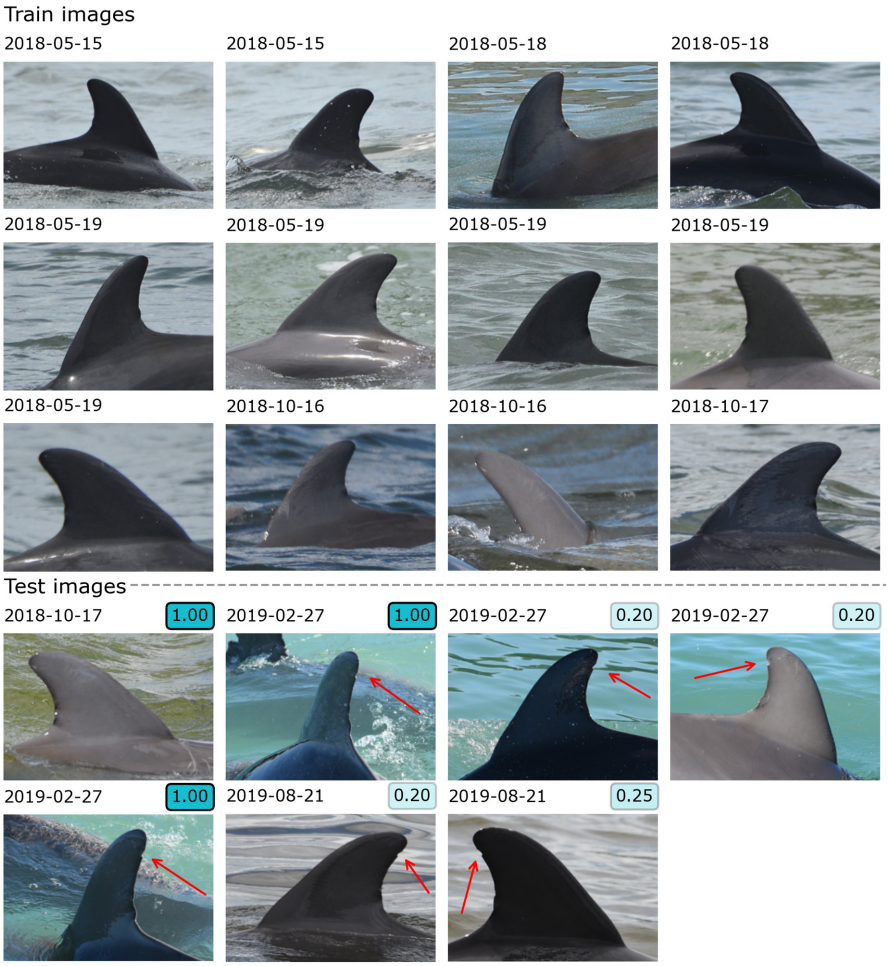

FIGURE 9 An example of a mark change and its influence on the model's predictions. The figure shows every train and test image for one individual bottlenose dolphin. The precision for the predicted identity in each test image is shown in a blue box. Note the degradation in prediction confidence before and after the new dorsal fin notch highlighted by a red arrow, occurred between 2018-10-17 and 2019-02-27.

GLOSSARY

Backbone	Convolutional neural networks designed to be widely applicable to many problems in image classification. Examples include ResNet, DenseNet, Xception and MobileNet (Khan et al., 2020).
Classes	Discrete output values in an image classification problem, for example rough-toothed dolphin (species) or HISb007 (identity).
Class imblance	Dataset characteristic whereby some classes have many images while others have very few. Often histograms of these datasets, showing the number of images by class, will be right-skewed.
Classification head (head)	Layer in the neural network that computes the classification probabilities. In logistic regression, the classification head is
Convolutional layer	Layer in a convolutional neural network that reduces the input, for example the raw RGB values of an image, using a kernel that extracts the input's component features, for example edges, textures or colours. The extracted features depend on the weights, which are learned during training.
Convolutional neural network (CNN)	A neural network consisting of convolutional layers and pooling layers. These networks have been shown to be effective at modelling images, audio and text (Khan et al., 2020).
Data augmentation	Image classification technique whereby images have been randomly altered, for example, converted to grayscale, mitigating the potential for overfitting.
Epoch	One pass through the full training dataset during training. Increasing the number of epochs leads to the model learning more about the training data risking the chance of overfitting.
Feature vector	A vector that summarizes an object such as an image, into component features. In image classification, the feature vector can be a, say, length 512 vector that summarizes a 1024 by 1024 by three RGB image. This summary is optimized during training to distinguish among classes, for example, identities. Thus, probability that two identities are distinct grows with the distance between their image's feature vectors.
Hyperparameter	Parameters that cannot be learned during training and therefore must be set a priori. Optimal values are typically chosen via validation.
Learning rate	Rate at which the optimizer changes values of the weights during training. In gradient descent, $\mathbf{w}_{t+1} = \mathbf{w}_t - \eta \frac{\partial \mathscr{L}}{\partial \mathbf{w}}$, where $\frac{\partial \mathscr{L}}{\partial \mathbf{w}}$ is the gradient of the loss with respect to the weights and η is the learning rate. A learning rate scheduler allows the learning rate to vary by epoch.
Pooling layer	Layer in a convolutional neural network that adds robustness to the model, for example robustness to changes in an object's location in an image. A pooling layer typically downsamples a convolutional kernel in the network, for example by simply outputting the maximum value (max pooling) of the kernel.
Neck	Any layers in the neural network that separate the backbone from the classification head(s) typically reducing the backbone's three dimensional output to a one dimensional feature vector.
Overfitting	Pitfall of machine learning, particularly for complex models. Overfit models will have high predictive performance on the training data but low predictive performance in general. In image classification, overfit models often memorize irrelevant portions of the training images, for example a boat in the background.
Validation	Process for finding optimal model structures or values of hyperparameters during which models are trained then evaluated on a hold out set. It can be distinguished from test, which is performed once at the end of development.
Transfer learning	Machine learning technique for applying learning from one problem to another. For examplein the shared architecture in Figure 2, the learning from the species classifier is shared with the identity classifier, improving the performance of the latter.
Weights	Model parameters that are learned during training.

4.1.8 | This algorithm is freely available

Finally, we note that this model is available as code in a GitHub repository and as a graphical user interface, with additional functionality for collaborative data management, via Happywhale.com.

While this model has been compared to the other 39,284 models in the competition, it has not been explicitly compared to cetacean photo–id models in the literature (e.g. Bergler et al., 2021; Maglietta et al., 2020, 2023; Thompson et al., 2021; Weideman et al., 2017). A full comparison (sensu Tyson Moore et al., 2022),

however, would be challenging because we are unaware of any model that could be reasonably used to predict the identities of every species in the dataset. Thus, any comparison would need to be piecemeal, comparing the state-of-the-art for each species to the full model presented here. Similarly, we recommend that future developers of multi–species models for cetaceans evaluate the performance of their models using the competition dataset, which has potential to be a powerful common dataset going forward.

We demonstrated that this approach can simultaneously achieve high predictive performance for classifying many species of cetaceans. We expect this approach will work well when applied to other multi–species catalogues if they arise from similar survey settings, for example camera trap surveys, and the species share identifying characteristics. With this in mind, we speculate that this multi–species approach would work with terrestrial mammals caught with camera traps, or individually identifiable animals from unoccupied aerial vehicle (UAV; drone) imagery, for example beluga whales, humpback whales, and crocodiles. Having reliable automation for these species and situations may make photo–id research more accessible to organizations, potentially quickening the pace of ecological and management-focused research.

AUTHOR CONTRIBUTIONS

Philip T. Patton, Taiki Yamaguchi, Kenshin Abe, Ted Cheeseman and Walter Reade conceived the ideas and designed methodology; Ted Cheeseman, Ken Southterland, Walter Reade, Addison Howard, Erin M. Oleson and Lars Bejder administered the project; Jason B. Allen, Erin Ashe, Aline Athayde, Robin W. Baird, Charla Basran, Elsa Cabrera, Júlio Cardoso, John Calambokidis, Emma L. Carroll, Amina Cesario, Barbara J. Cheney, Enrico Corsi, Jens Currie, John W. Durban, Erin A. Falcone, Holly Fearnbach, Kiirsten Flynn, Trish Franklin, Wally Franklin, Bárbara Galletti Vernazzani, Tilen Genov, Marie Hill, David Johnston, Erin L. Keene, Sabre D. Mahaffy, Tamara L. McGuire, Liah McPherson, Catherine Meyer, Robert Michaud, Anastasia Miliou, Dara N. Orbach, Heidi C. Pearson, Marianne H. Rasmussen, William J. Rayment, Caroline Rindaldi, Renato Rindaldi, Salvatore Siciliano, Stephanie Stack, Beatriz Tintore, Leigh G. Torres, Jared R. Towers, Cameron Trotter, Reny Tyson Moore, Caroline Weir, Rebecca Wellard, Randall Wells, Kymberly M. Yano and Jochen R Zaeschmar collected the data; Philip T. Patton, Taiki Yamaguchi and Kenshin Abe analysed the data; Philip T. Patton and Lars Bejder led the writing of the manuscript. All authors contributed critically to the drafts and gave final approval for publication.

AFFILIATIONS

[1]Marine Mammal Research Program, Hawai'i Institute of Marine Biology, University of Hawai'i at Mānoa, Kāne'ohe, Hawai'i, USA; [2]NOAA Fisheries Pacific Islands Fisheries Science Center, Honolulu, Hawai'i, USA; [3]Marine Ecological Research Centre, Southern Cross University, Lismore, New South Wales, Australia; [4]Happywhale.com, Santa Cruz, California, USA; [5]Preferred Networks, Inc., Chiyoda-ku, Tokyo, Japan; [6]Google, Kaggle, San Francisco, California, USA; [7]Chicago Zoological Society's Sarasota Dolphin Research Program, c/o Mote Marine Laboratory, Sarasota, Florida, USA; [8]Oceans Initiative, Seattle, Washington, USA; [9]Projeto Baleia à Vista (ProBaV), Ilhabela, Brazil; [10]Cascadia Research Collective, Olympia, Washington, USA; [11]Research Center in Húsavík, University of Iceland, Húsavík, Iceland; [12]Centro de Conservación Cetacea (CCC), Santiago, Chile; [13]School of Biological Sciences, University of Auckland-Waipapa Taumata Rau, Auckland, New Zealand; [14]Tethys Research Institute, Milan, Italy; [15]The Swire Institute of Marine Science, The University of Hong Kong, Hong Kong, Hong Kong; [16]School of Biological Sciences, University of Aberdeen, Cromarty, UK; [17]Pacific Whale Foundation, Wailuku, Hawai'i, USA; [18]SR3, SeaLife Response, Rehabilitation and Research, Des Moines, Washington, USA; [19]Marine Ecology and Telemetry Research, Seabeck, Washington, USA; [20]The Oceania Project, Hervey Bay, Queensland, Australia; [21]Morigenos-Slovenian Marine Mammal Society, Piran, Slovenia; [22]Sea Mammal Research Unit, Scottish Oceans Institute, University of St Andrews, St Andrews, UK; [23]Cooperative Institute for Marine and Atmospheric Research, Research Corporation of the University of Hawai'i, Honolulu, Hawai'i, USA; [24]Marine Science Department, Te Tari Putaiao Taimoana, University of Otago, Otago, New Zealand; [25]The Cook Inlet Beluga Whale Photo–ID Project, Anchorage, Alaska, USA; [26]School of Biological Sciences, Te Kura Mātauranga Koiora, University of Auckland, Auckland, New Zealand; [27]Groupe de Recherche et D'éducation sur les Mammifères Marins (GREMM), Tadoussac, Québec, Canada; [28]Archipelagos Institute of Marine Conservation, Samos Island, Greece; [29]Department of Life Sciences, Texas A&M University-Corpus Christi, Corpus Christi, Texas, USA; [30]Department of Natural Sciences, University of Alaska Southeast, Juneau, Alaska, USA; [31]Department of Marine Science-Te Tari Pūtaiao Taimoana, University of Otago, Dunedin, New Zealand; [32]L'association Evasion Tropicale, Bouillante, Guadeloupe; [33]Departamento de Ciências Biológicas, Escola Nacional de Saúde Pública/Fiocruz, Rio de Janeiro, Brazil; [34]Pacific Whale Foundation Australia, Urangan, Queensland, Australia; [35]Marine Mammal Institute, Oregon State University, Newport, Oregon, USA; [36]Bay Cetology, Alert Bay, British Columbia, Canada; [37]School of Engineering, Newcastle University, Newcastle, UK; [38]Falklands Conservation, Stanley, Falkland Islands; [39]Centre for Marine Science and Technology, Curtin University, Bentley, Western Australia, Australia; [40]Project ORCA, Perth, Western Australia, Australia; [41]Far Out Ocean Research Collective, Paihia, New Zealand and [42]Zoophysiology, Department of Bioscience, Aarhus University, Aarhus, Denmark

ACKNOWLEDGEMENTS

We thank the countless individuals who collected and/or processed the nearly 85,000 images used in this study and those who assisted, particularly those who sorted these images from the millions that did not end up in the catalogues. Additionally, we thank the other Kaggle competitors who helped develop the ideas, models and data used here, particularly those who released their datasets to the public. The graduate assistantship for Philip T. Patton was funded by the NOAA Fisheries QUEST Fellowship. This paper represents HIMB and SOEST contribution numbers 1932 and 11679, respectively. The technical support and advanced computing resources from University of Hawaii Information Technology Services—Cyberinfrastructure, funded in part by the National Science Foundation CC* awards # 2201428 and # 2232862 are gratefully acknowledged. Every photo–identification image was collected under permits according to relevant national guidelines, regulation and legislation.

CONFLICT OF INTEREST STATEMENT

We disclose no conflicts of interest.

PEER REVIEW

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/10.1111/2041-210X.14167.

DATA AVAILABILITY STATEMENT

The competition data are freely available at https://www.kaggle.com/competitions/happy-whale-and-dolphin. The data and code necessary to train, validate, and test the model are available at Zenodo (Abe, 2023) and GitHub https://github.com/knshnb/kaggle-happywhale-1st-place.

ORCID

Philip T. Patton https://orcid.org/0000-0003-2059-4355
Ted Cheeseman https://orcid.org/0000-0002-5805-2431
Erin M. Oleson https://orcid.org/0000-0002-4889-6059
Erin Ashe https://orcid.org/0000-0001-9918-1419
Aline Athayde https://orcid.org/0000-0001-6179-2943
Robin W. Baird https://orcid.org/0000-0002-9419-6336
Charla Basran https://orcid.org/0000-0002-6362-8474
John Calambokidis https://orcid.org/0000-0002-5028-7172
Júlio Cardoso https://orcid.org/0000-0003-0203-4493
Emma L. Carroll https://orcid.org/0000-0003-3193-7288
Amina Cesario https://orcid.org/0000-0003-3175-879X
Barbara J. Cheney https://orcid.org/0000-0003-4534-5582
Enrico Corsi https://orcid.org/0000-0001-7655-8754
Jens Currie https://orcid.org/0000-0001-6084-3091
Trish Franklin https://orcid.org/0000-0002-3253-4053
Wally Franklin https://orcid.org/0000-0003-3268-6775
Tilen Genov https://orcid.org/0000-0003-4814-8891
Sabre D. Mahaffy https://orcid.org/0000-0001-8255-192X
Anastasia Miliou https://orcid.org/0000-0002-7631-372X
Heidi C. Pearson https://orcid.org/0000-0003-0502-2105
Marianne H. Rasmussen https://orcid.org/0000-0002-6887-8616
Salvatore Siciliano https://orcid.org/0000-0002-0124-8070
Stephanie Stack https://orcid.org/0000-0002-7199-8408
Beatriz Tintore https://orcid.org/0000-0003-0615-2104
Jared R. Towers https://orcid.org/0000-0002-5700-1755
Reny Tyson Moore https://orcid.org/0000-0001-5541-193X
Caroline R. Weir https://orcid.org/0000-0002-2052-5037
Rebecca Wellard https://orcid.org/0000-0001-8427-564X
Randall Wells https://orcid.org/0000-0001-9793-4181
Kymberly M. Yano https://orcid.org/0000-0003-2054-3080
Jochen R. Zaeschmar https://orcid.org/0000-0002-6978-0995
Lars Bejder https://orcid.org/0000-0001-8138-8606

REFERENCES

Abe, K. (2023). Code from: A deep learning approach to photo-identification demonstrates high performance on two dozen cetacean species. *Zenodo*. https://doi.org/10.5281/zenodo.8010271

Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. (2019). Optuna: A next-generation hyperparameter optimization framework. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (pp. 2623–2631). Association for Computing Machinery.

Baird, R. W., Webster, D. L., Mahaffy, S. D., McSweeney, D. J., Schorr, G. S., & Ligon, A. D. (2008). Site fidelity and association patterns in a deep-water dolphin: Rough-toothed dolphins (*Steno bredanensis*) in the Hawaiian archipelago. *Marine Mammal Science*, *24*, 535–553. https://doi.org/10.1111/j.1748-7692.2008.00201.x

Bejder, L., Fletcher, D., & Bräger, S. (1998). A method for testing association patterns of social animals. *Animal Behaviour*, *56*, 719–725.

Bergler, C., Gebhard, A., Towers, J. R., Butyrev, L., Sutton, G. J., Shaw, T. J. H., Maier, A., & Nöth, E. (2021). Fin-print a fully-automated multi-stage deep-learning-based framework for the individual recognition of killer whales. *Scientific Reports*, *11*, 23480.

Bogucki, R., Cygan, M., Khan, C. B., Klimek, M., Milczek, J. K., & Mucha, M. (2019). Applying deep learning to right whale photo identification. *Conservation Biology*, *33*, 676–684.

Borchers, D., Distiller, G., Foster, R., Harmsen, B., & Milazzo, L. (2014). Continuous-time spatially explicit capture–recapture models, with an application to a jaguar camera-trap survey. *Methods in Ecology and Evolution*, *5*, 656–665.

Borowiec, M. L., Dikow, R. B., Frandsen, P. B., McKeeken, A., Valentini, G., & White, A. E. (2022). Deep learning as a tool for ecology and evolution. *Methods in Ecology and Evolution*, *13*, 1640–1660.

Cheeseman, T., Southerland, K., Park, J., Olio, M., Flynn, K., Calambokidis, J., Jones, L., Garrigue, C., Frisch Jordán, A., Howard, A., Reade, W., Neilson, J., Gabriele, C., & Clapham, P. (2021). Advanced image recognition: A fully automated, high-accuracy photo-identification matching system for humpback whales. *Mammalian Biology*, *102*, 1618–1476. https://doi.org/10.1007/s42991-021-00180-9

Clapham, M., Miller, E., Nguyen, M., & Darimont, C. T. (2020). Automated facial recognition for wildlife that lack unique markings: A deep learning approach for brown bears. *Ecology and Evolution*, *10*, 12883–12892.

Deng, J., Guo, J., Liu, T., Gong, M., & Zafeiriou, S. (2020). Sub-center arcface: Boosting face recognition by large-scale noisy web faces. In *European Conference on Computer Vision* (pp. 741–757). Springer.

Deng, J., Guo, J., Xue, N., & Zafeiriou, S. (2019). Arcface: Additive angular margin loss for deep face recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 4690–4699). The Computer Vision Foundation.

Ha, Q., Liu, B., Liu, F., & Liao, P. (2020). Google landmark recognition 2020 competition third place solution.

Hammond, P. S., Francis, T. B., Heinemann, D., Long, K. J., Moore, J. E., Punt, A. E., Reeves, R. R., Sepúlveda, M., Sigursson, G. M., Siple, M. C., Víkingsson, G., Wade, P. R., Williams, R., & Zerbini, A. N. (2021). Estimating the abundance of marine mammal populations. *Frontiers in Marine Science*, *8*, 1316.

Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In JMLR (Journal of Machine Learning Research) (Ed.), *International conference on machine learning* (pp. 448–456). PMLR.

Khan, A., Sohail, A., Zahoora, U., & Qureshi, A. S. (2020). A survey of the recent architectures of deep convolutional neural networks. *Artificial Intelligence Review*, *53*, 5455–5516.

Körschens, M., Barz, B., & Denzler, J. (2018). Towards automatic identification of elephants in the wild. *arXiv preprint arXiv:181204418*.

Loshchilov, I., & Hutter, F. (2016). SGDR: Stochastic gradient descent with warm restarts. *arXiv preprint arXiv:160803983*.

Loshchilov, I., & Hutter, F. (2017). Decoupled weight decay regularization. *arXiv preprint arXiv:171105101*.

Maglietta, R., Bussola, A., Carlucci, R., Fanizza, C., & Dimauro, G. (2023). Arianna: A novel deep learning-based system for fin contours analysis in individual recognition of dolphins. *Intelligent Systems with Applications*, *18*, 200207. https://doi.org/10.1016/j.iswa.2023.200207

Maglietta, R., Renò, V., Caccioppoli, R., Seller, E., Bellomo, S., Santacesaria, F. C., Colella, R., Cipriano, G., Stella, E., Hartman, K., Fanizza, C., Dimauro, G., & Carlucci, R. (2020). Convolutional neural networks

for Risso's dolphins identification. *IEEE Access*, *8*, 80195–80206. https://doi.org/10.1109/ACCESS.2020.2990427

Miele, V., Dussert, G., Spataro, B., Chamaillé-Jammes, S., Allainé, D., & Bonenfant, C. (2021). Revisiting animal photo-identification using deep metric learning and network analysis. *Methods in Ecology and Evolution*, *21*, 863–873.

Palencia, P., Fernández-López, J., Vicente, J., & Acevedo, P. (2021). Innovations in movement and behavioural ecology from camera traps: Day range as model parameter. *Methods in Ecology and Evolution*, *12*, 1201–1212.

Radenović, F., Tolias, G., & Chum, O. (2018). Fine-tuning cnn image retrieval with no human annotation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *41*, 1655–1668.

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2015). You only look once: Unified, real-time object detection.

Rosel, P., Mullin, K., Garrison, L., Schwacke, L., Adams, J., Balmer, B., Conn, P., Conroy, M., Eguchi, T., Gorgone, A., Hohn, A., Mazzoil, M., Schwartz, C., Sinclair, C., Speakman, T., Urian, K., Vollmer, N., Wade, P., Wells, R., & Zolman, E. (2011). Photo-identification capture-mark-recapture techniques for estimating abundance of bay, sound and estuary populations of bottlenose dolphins along the U.S. East Coast and Gulf of Mexico: A workshop report.

Tan, M., & Le, Q. (2019). Efficientnet: Rethinking model scaling for convolutional neural networks. In JMLR (Journal of Machine Learning Research) (Ed.), *International conference on machine learning* (pp. 6105–6114). PMLR.

Thompson, J. W., Zero, V. H., Schwacke, L. H., Speakman, T. R., Quigley, B. M., Morey, J. S., & McDonald, T. L. (2021). finFindR: Automated recognition and identification of marine mammal dorsal fins using residual convolutional neural networks. *Marine Mammal Science*, *38*, 139–150.

Tyne, J. A., Pollock, K. H., Johnston, D. W., & Bejder, L. (2014). Abundance and survival rates of the Hawai'i Island associated spinner dolphin (*Stenella longirostris*) stock. *PLoS ONE*, *9*, e86132.

Tyson Moore, R. B., Urian, K. W., Allen, J. B., Cush, C., Parham, J. R., Blount, D., Holmberg, J., Thompson, J. W., & Wells, R. S. (2022). Rise of the machines: Best practices and experimental evaluation of computer-assisted dorsal fin image matching systems for bottlenose dolphins. *Frontiers in Marine Science*, *9*. https://doi.org/10.3389/fmars.2022.849813

Urian, K., Gorgone, A., Read, A., Balmer, B., Wells, R. S., Berggren, P., Durban, J., Eguchi, T., Rayment, W., & Hammond, P. S. (2015). Recommendations for photo-identification methods used in capture-recapture models with cetaceans. *Marine Mammal Science*, *31*, 298–321. https://doi.org/10.1111/mms.12141

Weideman, H. J., Jablons, Z. M., Holmberg, J., Flynn, K., Calambokidis, J., Tyson, R. B., Allen, J. B., Wells, R. S., Hupman, K., Urian, K., & Stewart, C. V. (2017). Integral curvature representation and matching algorithms for identification of dolphins and whales.

Zhou, X., Girdhar, R., Joulin, A., Krähenbühl, P., & Misra, I. (2022). Detecting twenty-thousand classes using image-level supervision.

Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., Xiong, H., & He, Q. (2021). A comprehensive survey on transfer learning. *Proceedings of the IEEE*, *109*, 43–76. https://doi.org/10.1109/JPROC.2020.3004555

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

Supplementary S1. Characteristics for all 41 competition catalogues. MAP is the mean average precision on the test set. The two catalogues without a MAP score (8 and 21) were included in the competition yet omitted from this analysis. Train Img is the number of images in the training split. Test Img is the number of images in the test split. Train ID is the number of identities in the training split. Img / ID is the median number of images per identity in the training split. Img Width is the mean width of the images in the training split, in pixels. Box Width is a weighted mean of the bounding box size for images in the training split. A large difference in box width and image width implies that the images contain a lot of background noise.

Supplementary S2. Primary reference in the literature for each catalogue in the competition. Two catalogues (8 and 21) were included in the competition yet omitted from this analysis.

Supplementary S3. Figure S1: A schematic diagram of the two loss functions, inspired by Figure 3 in Deng et al. (2019). Each dorsal fin represents the image's projection onto an abstract, high-dimensional space (softmax loss) or a hypersphere (ArcFace loss). The three coloured boxes contain three images of a known individual dusky dolphin (*Lagenorhynchus obscurus*). Additionally, there is a tenth image of a "new individual," that is, one not in the training set. The white space between the boxes represents the decision boundaries. For softmax loss, the decision boundaries are so close that the identity of the new individual is ambiguous. In ArcFace, each sample has been pushed closer to its class centre, and away from the other centres, such that we can confidently classify the individual as new.

Table S1: Data augmentations used before training, implemented in the Albumentations library. translate_percent dictates the proportion of image that is translated. scale dictates the proportion of the image that is cropped, while ratio dictates the range of aspect ratio of the origin that is cropped. blur_limit dictates the max kernel size for blurring the image. grid is the size of the grid for splitting the image.

Table S2: Predictive performance by catalogue, including the mean average precision (MAP), Top1 Accuracy and number of images precision score.

How to cite this article: Patton, P. T., Cheeseman, T., Abe, K., Yamaguchi, T., Reade, W., Southerland, K., Howard, A., Oleson, E. M., Allen, J. B., Ashe, E., Athayde, A., Baird, R. W., Basran, C., Cabrera, E., Calambokidis, J., Cardoso, J., Carroll, E. L., Cesario, A., Cheney, B. J. … Bejder, L. (2023). A deep learning approach to photo–identification demonstrates high performance on two dozen cetacean species. *Methods in Ecology and Evolution*, *00*, 1–15. https://doi.org/10.1111/2041-210X.14167